Discussion of: “On Families of Distributions With Shape Parameters”

M. F. J. Steel* and F. J. Rubio†

We congratulate Prof. Jones on such an enjoyable and complete paper. The structured and organised format helps the reader gain understanding of a rather large and not always that transparent literature. We believe that this insightful survey which includes alternatives to the skew-symmetric family of distributions fills an important void in this area. As we are largely in agreement with the main points made in the paper, most of the discussion below is devoted to exploring slightly different perspectives (for example, induced by Bayesian inference) and adding detail on some of the issues mentioned in the paper. Section 2 presents some ideas for combining two of the families of distributions discussed in the paper into a new family, with potentially interesting properties.

1 Desiderata for flexible univariate continuous distributions

We agree with the desirable properties that are proposed for families of distributions in the Introduction of the paper. Rubio (2013) presents a discussion on this point in the context of distributions for modelling skewness and kurtosis. However, his list contains an additional desirable property, namely “flexibility with respect to an interpretable measure of skewness and/or kurtosis”. The interpretation of the shape parameters of a distribution as skewness or kurtosis parameters typically derives from the partial ordering proposed by van Zwet (1964), or simply from noting that by varying the parameter values one can control the tailweight and the asymmetry of the density function. However, it is not common practice to focus on the range of skewness or kurtosis values that can be obtained by varying the shape parameters. If anything, such examinations are conducted as an afterthought, but the consideration of how much skewness or kurtosis can be generated within a family of distributions does not usually form part of the primary desiderata used in selecting a distribution. In the context of distributions obtained by adding shape parameters to a symmetric distribution, it is important not only to analyse the role of the shape parameters, but also to quantify the flexibility induced by these new parameters. This could also help practitioners in selecting a model in terms of the amount of skewness or kurtosis required in their context. In particular, they might want to rule out some methods for adding parameters that do not induce sufficient flexibility. For example, as mentioned in the paper, the Marshall and Olkin (1997) transformation is not something we would recommend as a skewing mechanism on the basis of these considerations.

*University of Warwick, Department of Statistics, Coventry, CV4 7AL, UK. E-mail: M.F.Steel@stats.warwick.ac.uk

†University of Warwick, Department of Statistics, Coventry, CV4 7AL, UK. E-mail: Francisco.Rubio@warwick.ac.uk

Of course, an important challenge in this context is to choose an "interpretable" (that word, once again!) measure of kurtosis or skewness. Most authors opt for the use of the standardised third central moment for measuring skewness. This quantity satisfies the required properties to be considered a measure of skewness (Arnold and Groeneveld, 1995), but it is difficult to interpret (and it requires the existence of the third moment, which restricts its applicability). The reason for this difficulty is that this measure is mostly driven by the tails, implying that two distributions with the same standardised third central moment may have very different levels of asymmetry. Figure 1 shows two density functions (a two-piece normal density and a skew-symmetric t density with 4 degrees of freedom) with the same standardised third central moment (0.9) but arguably very different levels of asymmetry, with the skew-symmetric t appearing virtually symmetric. Alternative (and more interpretable) scalar and functional measures of asymmetry and kurtosis are proposed and discussed in e.g. Critchley and Jones (2008).

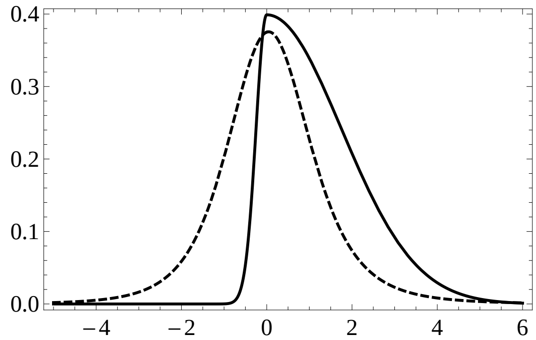


Figure 1: Two-piece normal (continuous line) and Skew-symmetric t with 4 degrees of freedom (dashed line), possessing the same standardised third central moment.

The third requirement mentioned in the paper, namely interpretability of the roles of the parameters, becomes even more important if we consider Bayesian methods of inference. It is, in particular, critical for the specification of the prior. If parameters have clearly defined roles, we typically find it much easier to elicit reasonable priors, as it is often feasible to think of reasonable priors on interpretable measures of skewness or kurtosis and this then induces priors on the skewness or kurtosis parameters (if these measures are injective functions of the corresponding parameters). This also leads to very natural mechanisms for the matching of priors in different models (for which the shape parameters are model-specific), so that formal Bayesian model comparison through Bayes factors can be conducted meaningfully. In addition, for clearly interpretable parameters it is much more realistic to assume independent or product structure priors. If various parameters jointly have a substantial impact on more than one aspect of the distribution, all these issues become much harder, and sensible prior elicitation strategies are hard to design in these cases.

2 Combining families of distributions

In terms of methods used for generating new distributions, the paper mentions two general representations which can be classified as transformations applied to the distribution function or generalised probability integral transformations (Family 4, Ferreira and Steel, 2006), and parametric transformations of random variables (Family 2, Ley and Paindaveine, 2010). These two approaches have complementary advantages. By using the representation in Ley and Paindaveine (2010) it becomes natural to add kurtosis parameters, since this method has a direct link with the definition of kurtosis in van Zwet (1964); the representation in Ferreira and Steel (2006) can be easily used to induce asymmetry while fixing the mode. The evaluation of the mode of the distributions generated through Ley and Paindaveine (2010) is typically more challenging than for those using Ferreira and Steel (2006). This makes the combination of the two representations tempting, which can be easily done as follows.

Result 1 *Let G be a symmetric unimodal distribution from the location-scale family, $P : [0,1] \to [0,1]$ be a distribution function, and $H : \mathbb{R} \to \mathbb{R}$ be an increasing diffeomorphism. Then, $F(x) = P\{G[H(x)]\}$ is also a distribution function.*

Thus, we would expect that the combined representation presented in Result 1 can be used to produce flexible distributions with easily interpreted parameters. As an illustrative example of the types of distributions we can generate in this way, we combine the symmetric sinh-arcsinh transformation $H(x;\delta) = \sinh[\delta \operatorname{arcsinh}(x)]$ (equation (4) in the paper with $a = 0$), which includes a shape parameter δ that controls kurtosis, and the two-piece transformation. The latter is parameterized as in Mudholkar and Hutson, 2000, which leads to (7) in the paper as a scale transformation, but, as explained in Ferreira and Steel, 2006 this can also be generated as a member of Family 4. More specifically, for a symmetric unimodal distribution S with density s, it corresponds to a choice for P with density function (denoted by w' in (8) in the paper)

$$p(u;\gamma) = \frac{s\left[\frac{1}{1-\text{sign}\{u-(1/2)\}\gamma} S^{-1}(u)\right]}{s[S^{-1}(u)]},$$

which is a useful mechanism to induce skewness, depending on a skewness parameter $\gamma \in (-1,1)$. In our example, it is applied to $S(x;\delta) = G[H(x;\delta)]$. It preserves the mode (see Theorem 2 in Ferreira and Steel, 2006) and, as γ varies in $(-1,1)$ it can cover the entire range of the Arnold and Groeneveld (1995) measure of skewness, which simply equals $-\gamma$.

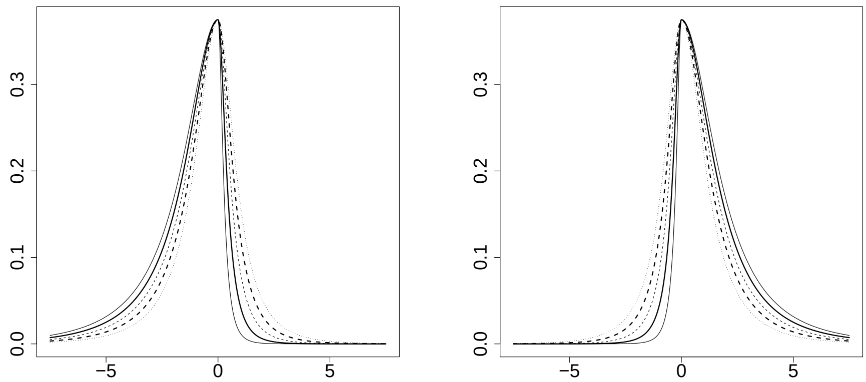
We apply this combination to the hyperbolic secant distribution as the symmetric unimodal distribution G. The resulting density can be written in closed form as follows:

$$f(x;\delta,\gamma) = \frac{\delta}{2} C(x,\delta,\gamma) \operatorname{sech}\left\{\frac{\pi}{2}\sinh\left[\delta \operatorname{arcsinh}\left(\frac{x}{1-\text{sign}\{x\}\gamma}\right)\right]\right\}, \tag{1}$$

where $C(x,\delta,\gamma) = \cosh\left[\delta \operatorname{arcsinh}\left(\frac{x}{1-\text{sign}\{x\}\gamma}\right)\right]\left\{1+\left(\frac{x}{1-\text{sign}\{x\}\gamma}\right)^2\right\}^{-1/2}$.

Figure 2 shows some examples of the shapes we can obtain with this distribution, which could be called the two-piece hyperbolic secant sinh-arcsinh (TP Sech-SAS) distribution. By construction, the density has mode at 0 and contains two parameters (δ,γ) which separately control the tails and skewness, respectively. As mentioned above, by varying γ, this distribution can cover the whole range of the measure of skewness in Arnold and Groeneveld (1995), in contrast to the HS-SAS distribution proposed by Fischer and Herrmann (2013). The latter distribution differs from (1) only in the way skewness is introduced. The tails in (1) are controlled by δ, where $\delta > 1$ leads to heavier tails than the hyperbolic secant distribution (which has excess kurtosis of 2) and $\delta < 1$ generates lighter tails. All moments of the distribution in (1) will exist; this follows from the proof in Fischer and Herrmann (2013) together with the fact that moment existence is not affected by the two-piece transformation (Ferreira and Steel, 2006).

Another choice for $P(\cdot, \gamma)$ could be the Beta distribution with shape parameters $(\gamma, 1/\gamma)$ (see Ferreira and Steel, 2006), although this transformation may not induce a lot of flexibility. A more complete analysis of the flexibility of these models would require the use and comparison of other measures of asymmetry and kurtosis.



(a) (b)

Figure 2: TP Sech-SAS densities: (a) left-asymmetric examples, and (b) right-asymmetric examples.

3 Main body versus tails

We can distinguish two types of asymmetry of a distribution, which Prof. Jones characterises as "main-body-skewness" and "tail-skewness". It is helpful for the discussion below to define the following:

Definition 1 *Let F be the distribution of a random variable X in $\Re$ with probability density function f. We say that the tail behaviour in each direction is the same if and only if there exists $K > 0$ such that $\lim_{x\to\infty} \frac{f(x)}{f(-Kx)}$ is positive and finite.*

In general, we can classify unimodal asymmetric distributions into the following two classes:

Class 1: Asymmetric distributions with the same tail behaviour in each direction.

Class 2: Asymmetric distributions with different tail behaviour in each direction.

Roughly speaking, distributions from Class 1 exhibit "main-body-skewness", while distributions from Class 2 also exhibit "tail-skewness". Class 1 contains the family of two-piece distributions (including the one introduced in (1) in Section 2 of this comment) and the members of Family 4 for which the skewing mechanism w' does not tend to either zero or infinity at the boundaries of the unit interval (see Theorem 3 of Ferreira and Steel, 2006). Most of the distributions mentioned in the paper belong to Class 2. By choosing a distribution from either Class 1 or Class 2 for modelling, we are implicitly making assumptions on the type of asymmetry of the data. This can be a limitation if one is trying to understand the nature of the asymmetry. In order to come up with families of distributions that can capture (in a controlled and interpretable manner) both types of asymmetry, Rubio and Steel (2014) proposed the family of double two-piece (DTP) distributions, which generalises the construction in Zhu and Galbraith (2010). Rubio and Steel (2014) distinguish two subfamilies, two-piece scale (TPSC) and two-piece shape (TPSH) distributions, that can capture skewness in the main body and the tails, respectively. By using the DTP family of distributions, we do not need to make strong assumptions about the type of asymmetry in the data and, by conducting formal model comparison, we can learn about the type of asymmetry favoured by the data. Other 5-parameter distributions, although flexible, often do not seem useful for identifying the type of asymmetry.

4 Semiparametric distributions

The paper briefly mentions semiparametric versions of distributions in Families 1 and 2. In the context of Family 4 a similar semiparametric approach is examined in Quintana et al. (2009), who propose flexible representations for either w' or g in (8) of the paper. As a first class of flexible distributions (which they denote by "Bernstein-skew"), they propose the use of Bernstein densities (a weighted mixture of specific Beta densities; see e.g. Petrone and Wasserman, 2002) to model the skewing mechanism w'. This makes the tails in both directions equal to the tails of the underlying symmetric distribution G. It is only by restricting the weights of the most extreme Beta densities to zero that we can make the tails thinner than those of G (if, for example, G is Student t, we then increase the moment existence in either tail by integer multiples).

For the second class of distributions, they start with a general unimodal symmetric distribution, and incorporate a skewing mechanism w' that is constrained to be unimodal with mode at $1/2$. From Theorem 2 in Ferreira and Steel (2006), this ensures that the resulting skewed distribution is unimodal with the same mode as g, which is desirable property (i) in the paper and is necessary in the context of modal regression models. In particular, Quintana et al. (2009) choose the constructed skewing mechanism with proportional tails of Ferreira and Steel (2006). This particular choice accommodates skewness around the mode and in the tails of the distribution, without altering moment existence (in the terminology of Definition 1 above, the tail behaviour is the same in each direction). This approach maintains unimodality and controls skewness, since the Arnold and Groeneveld (1995) measure of skewness is only a function of the parameters of w' and does not depend on g. For G Quintana et al. (2009) use the representation of unimodal symmetric densities as mixtures of uniform distributions

$$g(x) = \int_0^\infty \frac{1}{\theta} I\{x \in (-\theta, \theta)\} dQ(\theta),$$

where $I\{\cdot\}$ denotes the indicator function and where Q is a distribution function on the positive real line. To flexibly model G, they propose to adopt a stick-breaking prior distribution with a finite number of terms for Q, which is a commonly used random probability measure.

References

Arnold, B. C. & Groeneveld, R. A. (1995). Measuring skewness with respect to the mode. *Amer. Statist.*, **49**, 34–38.

Critchley, F. & Jones, M. C. (2008). Asymmetry and gradient asymmetry functions: density-based skewness and kurtosis. *Scand. J. Stat.*, **35**, 415-437.

Ferreira, J. T. A. S. & Steel, M. F. J. (2006). A constructive representation of univariate skewed distributions. *J. Amer. Statist. Assoc.*, **101**, 823–829.

Fischer, M. & Herrmann, K. (2013). The HS-SAS and GSH-SAS Distribution as Model for Unconditional and Conditional Return Distributions. *Austrian J. Stat.*, **42**, 33–45.

Jones, M. C. (2004). Families of distributions arising from distributions of order statistics. *TEST*, **13**, 1–43.

Ley, C. & Paindaveine, D. (2010). Multivariate skewing mechanisms: A unified perspective based on the transformation approach. *Stat. Probabil. Lett.*, **80**, 1685–1694.

Marshall, A. W. & Olkin, I. (1997). A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families. *Biometrika*, **84**, 641–652.

Petrone, S. & Wasserman, L. (2002). Consistency of Bernstein polynomial posteriors. *J. R. Stat. Soc. Ser. B*, **64**, 79–100.

Mudholkar, G. S. & Hutson, A. D. (2000). The epsilon-skew-normal distribution for analyzing near-normal data. *J. Stat. Plan. Infer.*, **83**, 291–309.

Quintana, F. A., Steel, M. F. J. & Ferreira, J. T. A. S. (2009). Flexible univariate continuous distributions. *Bayesian Anal.* **4**, 497–522.

Rubio, F. J. (2013). *Modelling of Kurtosis and Skewness: Bayesian Inference and Distribution Theory.* PhD Thesis, University of Warwick, UK.

Rubio, F. J. & Steel, M. F. J. (2014). Bayesian modelling of skewness and kurtosis with two-piece scale and shape transformations. arXiv:1406.7674.

van Zwet, W. R. (1964). *Convex Transformations of Random Variables.* Mathematisch Centrum, Amsterdam.

Zhu, D. & Galbraith, J. W. (2010). A generalized asymmetric Student-t distribution with application to financial econometrics. *J. Econometrics*, **157**, 297–305.